PUBLIC TECHNICAL WHITE PAPER / V1.2

Organizational Intent as Infrastructure for Enterprise AI

A technical validation and public architecture brief for governed decision context

CORE FINDING

In a sealed multi-model holdout benchmark, Nucadia achieved 92.8% action correctness and 95.6% behavior correctness. It substantially outperformed raw and conventional context, while performing comparably to a deliberately strong Ideal Expert Brief baseline.

Public disclosure note. This paper describes the problem, validation design, benchmark results, product value, and high-level architecture, including the control-plane and capability model now covered by the filed provisional application. Proprietary resolution logic, schemas, precedence rules, binding logic, compiler behavior, and implementation details are intentionally omitted.

Brand note. Nucadia is the current public brand. The provisional patent application and sealed V2.1 benchmark were prepared under the earlier working name IntentSpec. The underlying architecture, mechanism, and benchmark condition are unchanged.

Nucadia

Technical validation brief | September 2026

Executive summary

Enterprise AI systems are becoming increasingly capable at retrieving information, generating content, and taking action. A persistent problem remains: access to data does not reliably tell an AI system what the organization is trying to accomplish, which objectives take priority, which constraints apply, or how the current task should be interpreted in context.

Nucadia is designed as a governed intent and context control plane for enterprise AI. It supports AI-assisted authoring of machine-readable organizational intent, task requirements, context relationships, source-authority rules, and capability semantics; separates proposed configuration from attested or active configuration; and applies the governed specification at runtime to a specific actor, task, situation, current state, and externally authorized capability envelope. The runtime output is a versioned Action Context that can carry active intent, selected context, constraints, prerequisites, coverage state, provenance, and capability applicability to downstream AI systems and agents. The exact resolution mechanism remains proprietary and is not described in this public paper.

A sealed V2.1 holdout benchmark, conducted under the earlier working name IntentSpec, tested whether this approach could improve AI decision quality on unseen cases without relying on a hand-authored answer key in the prompt. The benchmark used three acting model families, non-self judging, three repetitions, paired counterfactual scenarios, permission-sensitive cases, null cases, and conflict cases.

Condition	Description	Action correctness	Behavior correctness
A	Raw task context	41.9%	53.7%
B	Conventional production context	47.5%	70.4%
C	Ideal Expert Brief	91.7%	94.2%
D	**Nucadia decision context***	**92.8%**	**95.6%**

Nucadia improved action correctness by 50.9 percentage points versus raw context and 45.4 points versus conventional context. Against the Ideal Expert Brief, Nucadia was effectively at parity on the primary outcome measures. The result supports the core mechanism hypothesis, but it does not establish production security, enterprise adoption, economic value, or real-world business outcomes.

What this paper claims

Structured organizational intent can materially change AI decision quality on tasks where goals, priorities, constraints, or permissions determine the right action.

A system-generated decision context can approach the quality of an ideal manually constructed expert brief in a synthetic holdout benchmark.

The benchmark supports technical mechanism validation. It does not yet constitute platform validation or market validation.

The broader product architecture extends the validated context mechanism into a governed control plane that can author and maintain intent, represent task and capability semantics, and resolve decision-specific Action Context for downstream AI. Capability-aware orchestration and production economics remain product hypotheses rather than benchmark-validated claims.

1. The missing layer in enterprise AI

Most enterprise AI stacks are optimized around model capability, tool access, retrieval, and workflow automation. Those layers help a system know what information exists and what actions are possible. They do not, by themselves, provide a durable representation of what the organization currently intends.

This distinction matters because the same facts can support different correct actions under different strategies. A customer success team may prioritize retention in one period and margin discipline in another. An operations team may favor throughput until safety or quality constraints become binding. A recruiting team may pursue speed while still being prohibited from using protected attributes. The information available to the AI can remain largely unchanged while the correct decision changes.

DESIGN THESIS

Enterprise AI needs more than retrieval. It needs a reliable way to interpret facts relative to current goals, authority, constraints, and user intent at the moment of action.

2. Public architecture

Nucadia is model-agnostic infrastructure with two related planes. In the control plane, enterprise sources such as strategy, policy, role definitions, workflows, schemas, data catalogs, and tool descriptions can be converted into proposed machine-readable configuration, then reviewed, attested, versioned, and activated under organizational authority. In the runtime plane, the Resolver applies the active governed specification to a particular actor, task, situation, current state, and externally authorized capability envelope to produce a decision-specific Action Context for an AI system, agent, workflow engine, or other computing process.

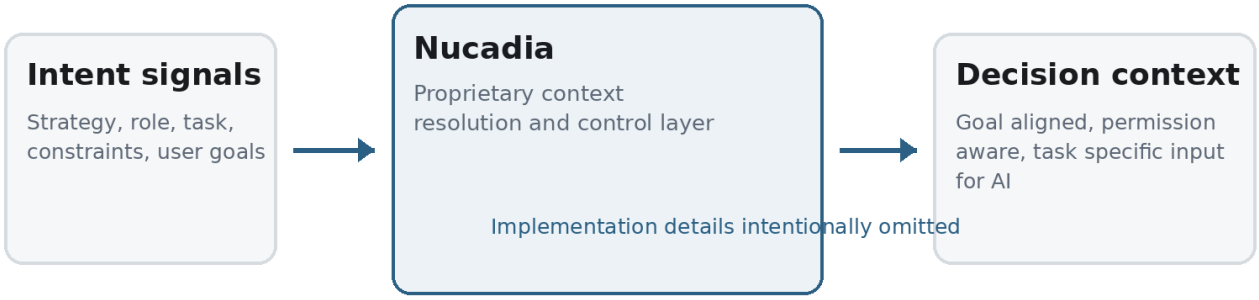

Figure 1. Simplified public architecture. The proprietary resolution mechanism remains intentionally abstracted.

Publicly described	Kept proprietary
AI-assisted authoring can propose structured intent, task requirements, context relationships, source-authority rules, and capability semantics from enterprise sources.	Exact extraction prompts, schema-generation logic, binding logic, and proposal-ranking methods.
Proposed configuration can remain non-authoritative until reviewed, attested, imported from an authoritative source, or activated by a governed rule.	Precedence, inheritance, override, source-authority scoring, and conflict-resolution algorithms.
Runtime resolution applies governed configuration to an actor, task, situation, current state, and externally authorized capability envelope.	Exact resolver algorithms, predicate implementation, optimization logic, and compiler behavior.
The output is a versioned Action Context that can contain active intent, selected context, unknowns, prerequisites, coverage state, provenance, and applicable capabilities.	Exact schemas, normalization logic, internal reason-code structure, and model-specific transformations.
RAG and other retrieval systems can supply candidate context; external IAM or policy systems remain authoritative for capability access.	Internal context-selection heuristics, side-channel controls, cache strategy, and proprietary maintenance algorithms.

Author, govern, resolve

The control plane is designed around a separation of duties: AI can help create and maintain the specification, but model-generated proposals do not become organizational truth merely because a model produced them. An authorized person, system of record, trusted service, or configured governance rule can attest or activate configuration. At runtime, the Resolver primarily applies the governed configuration; model assistance can be used for bounded tasks such as natural-language task mapping, entity extraction, candidate retrieval, semantic ranking, or ambiguity detection.

Action Context as the runtime contract

The Action Context is the machine-readable runtime artifact supplied to downstream AI. Publicly described fields can include active intent, mandatory constraints, selected context, unknowns, blocking or nonblocking prerequisites, coverage states, provenance, source timestamps, intent and resolver versions, capability applicability, confidence or sufficiency state, expiry, and re-resolution conditions. This structure supports traceability and replay without requiring every downstream model to reconstruct organizational meaning from raw documents.

Capabilities and agent planning

Nucadia can represent a canonical capability independently from a specific vendor tool or API implementation. A capability can carry semantic metadata such as best-use cases, poor-fit conditions, prerequisites, inputs, effects, risk, cost, reversibility, dependencies, provenance, and version. At runtime, the Resolver can evaluate capability applicability only within a capability envelope established by external identity, policy, entitlement, or application systems. A downstream agent can then dynamically plan and sequence among applicable capabilities, with re-resolution after tool results or other state changes.

Retrieval is an input, not the governing layer

Retrieval-augmented generation, semantic search, vector retrieval, keyword search, graph traversal, databases, APIs, and event streams can all provide candidate context. Nucadia is intended to determine how candidate information should matter relative to active intent, task requirements, authority, freshness, constraints, prerequisites, and current state. Retrieval therefore can remain a useful subordinate component rather than being replaced.

3. Validation design

* Benchmark tables use the Nucadia brand for readability; the sealed run itself was executed under the working name IntentSpec.

The V2.1 holdout was built to test a narrow technical question: can Nucadia reliably produce expert-quality decision context on unseen cases without obvious governance or exposure failure? The experiment deliberately included a strong expert baseline so that success did not depend on beating weak prompting.

Element	Holdout design
Scenario set	24 sealed counterfactual pairs, represented as 48 task instances.
Acting models	Three model families: OpenAI, Anthropic, and Google.
Conditions	A raw context, B conventional context, C Ideal Expert Brief, D Nucadia*.
Repetitions	Three repetitions for each actor, instance, and condition.
Acting coverage	1,728 of 1,728 valid actor responses.
Judging	864 of 864 valid, non-self, unambiguous judgments.
Scenario types	Reversal, null, conflict or missing-context, and permission-sensitive cases.
Primary comparison	D versus A, B, and the strong C baseline.

Why the Ideal Expert Brief matters

Condition C was intentionally difficult to beat. It represented a high-quality brief that assumed an expert had access to the correct permitted organizational, role, task, and situational context for the decision. This is an economically unrealistic operating model at scale, but it is a useful quality ceiling. Nucadia should therefore be interpreted as a systematic path to expert-quality contextualization, not as evidence that automation is inherently better than experts.

Counterfactual structure

Paired scenarios were designed so that the visible operating facts could remain similar while a relevant objective, constraint, permission, or situational condition changed. This structure makes it possible to test whether an AI system changes behavior when the governing intent changes, while remaining stable when the changed information should not affect the decision.

4. Results

Primary outcomes

Metric	A Raw	B Conventional	C Expert brief	D Nucadia*

Action correctness	41.9%	47.5%	91.7%	92.8%
Behavior correctness	53.7%	70.4%	94.2%	95.6%
Critical failure rate	2.08%	3.01%	0.69%	0.69%
Permission compliance	100%	100%	100%	100%
Unauthorized context exposure detected	0	0	0	0

The primary result is the separation between A and B versus C and D. Better generic prompting improved behavior, but it did not close the gap created by missing intent. The condition now represented by Nucadia matched or slightly exceeded the expert baseline on aggregate action and behavior correctness.

Source: sealed V2.1 holdout run 20260916_021820_900430, conducted under the earlier working name IntentSpec.

Intent-sensitive behavior

Metric	A Raw	B Conventional	C Expert brief	D Nucadia*
Conflict clarification accuracy	52.2%	46.7%	90.0%	95.6%
Reversal accuracy	1.2%	0.0%	96.3%	97.5%
Null stability	91.1%	86.7%	100%	100%

Reversal accuracy is especially diagnostic. In reversal cases, the correct action changes because the governing intent or relevant condition changes. A and B almost never reversed correctly. C and D did so reliably. At the same time, D preserved 100% null stability, indicating that it did not simply become more reactive to every change.

Independent judge comparisons

Comparison	Mean score delta	Wins	Ties	Losses
D vs A	+10.91	396	28	8
D vs B	+10.80	415	10	7
D vs C	+0.19	92	201	139

The D versus C result should be read as parity, not dominance. C won more individual judge comparisons than D, while D had a slightly positive mean score delta overall. The preregistered unit-level non-inferiority criterion was met in 393 of 432 units, or 91.0%, above the 80% acceptance threshold.

Resolver coverage and exposure

Measure	Result
Eligible universe evaluated	432 / 432
Context records included	1,206
Unavailable context states	72

Unauthorized protected counts exposed	None
Detected ground-truth answer leakage	None

These checks are important, but they are not a production security certification. The benchmark tested for explicit unauthorized context exposure and answer leakage within the experimental environment. It cannot reveal covert reasoning and does not substitute for enterprise security engineering, access-control review, or red-team testing.

5. Value proposition

The benchmark suggests that the value of Nucadia is not another general-purpose prompt template. Its potential value is to make organizational intent and decision context reusable infrastructure across AI systems and tasks. The broader product architecture adds governed authoring and maintenance, task and coverage semantics, capability-aware planning support, provenance, and versioned runtime resolution. These broader platform benefits are product hypotheses until tested in production.

Enterprise problem	Nucadia value hypothesis
AI has data but lacks strategic direction.	Translate current organizational intent into task-specific decision context.
Prompt quality depends on individual user skill.	Reduce dependence on each user knowing which goals, rules, and constraints to restate.
Strategy changes faster than prompts and workflows.	Create a control point where changes can be reflected systematically across AI use cases.
Too much context increases noise and risk.	Supply context that is relevant to the decision rather than everything the system can access.
Multiple AI vendors create inconsistent alignment.	Provide a model-agnostic layer that can sit upstream of different models and agents.
Governance is often separated from usefulness.	Treat authority, context, and intent as part of the same decision-context problem.
Agents have many authorized tools but weak guidance about which fit the current objective.	Resolve applicability among already-authorized capabilities while leaving IAM, credentials, and execution with existing systems.

AI systems cannot easily explain why a particular fact, constraint, or version governed a decision.	Carry provenance, source state, coverage semantics, versions, and re-resolution conditions in the Action Context.
Repeated runtime interpretation can duplicate prompts, policy reading, and tool-schema payloads.	Move stable organizational semantics into governed configuration and potentially narrow runtime context and capability sets; production cost and energy effects remain unvalidated.

Who this is for

The near-term audience is enterprise AI platform teams, agent builders, AI product leaders, architects, and governance teams that are moving beyond isolated chat assistants toward systems that make or recommend consequential decisions. The strongest fit is likely to be environments where goals change, permissions matter, multiple roles interact, and a single generic prompt is insufficient.

6. Implications for enterprise AI

Model capability is not the same as organizational alignment. More capable models still need a reliable representation of what matters for the current decision.

Retrieval and intent are complementary. Retrieval answers what information is relevant to a query. Intent determines how that information should be interpreted relative to goals and authority.

Expert briefing quality can become infrastructure. The benchmark suggests that high-quality contextualization can be systematized rather than reconstructed manually for every user and task.

Strategy can become executable. If organizational intent is represented as infrastructure, a change in strategy can potentially affect downstream AI behavior without manually rewriting every prompt.

Governance can move closer to the moment of action. Permission-aware context selection creates a path to make usefulness and control part of the same runtime context decision.

MOST IMPORTANT IMPLICATION

The central enterprise AI problem may not be access to more information. It may be ensuring that each AI system receives the right organizational meaning for the decision it is making.

Action Context can become an auditable runtime contract. Version, provenance, coverage, prerequisite, and re-resolution metadata can make it easier to reconstruct what information and intent governed a particular AI-assisted action.

Governed authoring can amortize semantic interpretation. AI can assist with extracting and proposing organizational meaning during control-plane maintenance, while runtime resolution can rely primarily on reviewed configuration. Whether this materially reduces production cost, latency, tokens, or energy remains to be measured.

Capability semantics can become infrastructure. Separating what an organization can do from the particular tool that implements it gives agents a more stable basis for planning across changing vendors and interfaces.

7. What remains unproven

This work validates a mechanism in a synthetic benchmark. It should not be interpreted as evidence that a finished platform is ready for enterprise deployment or that customers will purchase it.

The benchmark does not prove that real enterprise systems can maintain intent accurately at organizational scale.

It does not establish implementation cost, latency, reliability, or total cost of ownership in production.

It does not prove that the approach improves revenue, productivity, customer outcomes, or risk outcomes in live organizations.

It does not validate integrations, identity systems, access controls, network security, auditability, or compliance requirements.

AI-based judging is not identical to human expert judgment. Human evaluation remains useful for external validation.

Token counts in the benchmark are estimates based on characters.

Exposure detection is limited to explicit values and distinctive strings. Behavioral compliance cannot reveal covert reasoning.

Potential reductions in input tokens, retrieval volume, tool-schema payloads, model calls, cost, latency, or energy use have not been validated in production. Token count is not a direct measure of energy consumption.

The benchmark does not validate capability-aware tool narrowing, end-to-end agent planning, or dynamic tool sequencing. Those are product hypotheses supported by the same resolution architecture and require separate testing.

8. Disclosure boundary

This public paper intentionally stops at the system boundary. It publicly describes the control-plane lifecycle, AI-assisted authoring and attestation, task and capability concepts, high-level Action Context contents, the external authorization boundary, re-resolution, and the relationship to retrieval. The exact representation of intent, binding logic, authority resolution, precedence rules, change

propagation algorithms, compiler behavior, context coverage implementation, ranking heuristics, and evaluation internals remain proprietary.

This disclosure strategy is deliberate. Nucadia is early enough that the implementation itself remains an important source of defensibility. Future product documentation may describe public interfaces and developer contracts without disclosing the internal resolution mechanism.

9. Conclusion

The V2.1 sealed holdout supports a focused conclusion: decision-specific organizational intent materially improved AI decision quality on unseen tasks, and the condition now represented by Nucadia performed comparably to an ideal manually constructed expert brief while preserving strong reversal behavior, conflict handling, null stability, and explicit permission compliance.

The next meaningful validation step is not another synthetic benchmark. It is product and market validation with real enterprise workflows, real source systems, human experts, and design partners. That validation should test not only decision quality but also maintainability, integration burden, latency, auditability, capability selection, re-planning, token and model-call efficiency, and the economics of keeping organizational intent current.

PUBLIC CLAIM

Nucadia is a governed intent and context control plane for enterprise AI. It helps organizations author and maintain machine-readable intent, task requirements, context relationships, source authority, and capability semantics, then resolves what applies for a specific actor, task, situation, and externally authorized capability envelope into a versioned Action Context for downstream AI systems and agents. In a sealed synthetic multi-model holdout, the validated decision-context mechanism achieved 92.8% action correctness versus 41.9% to 47.5% for raw or conventional context and performed comparably to an ideal expert brief at 91.7%.

Nucadia

Organizational intent as governed infrastructure for AI